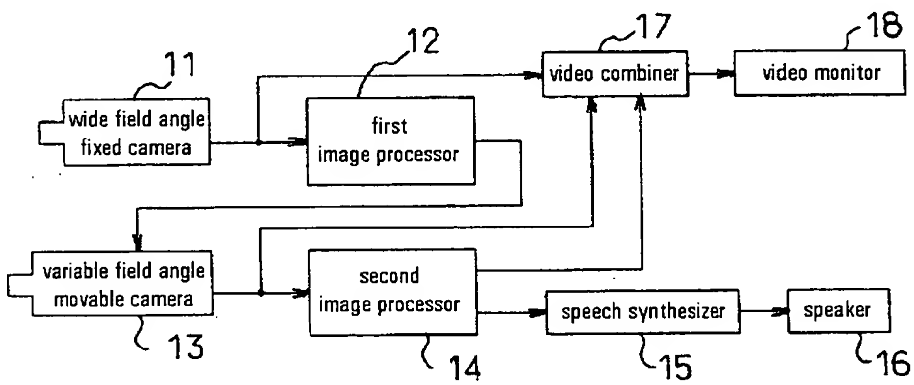
Translation of Japanese Unexamined Pat. Appl. Publication No. 06-225197

Application No.	03-178099
Filing Date	18 July 1991
Publication No.	06-255197
Publication Date	12 August 1994
Int. Cl.[5]	H04N 5/232; G06F 15/70; G08B 13/196
Inventor	Isamu YOROISAWA
Applicant	NTT Corporation

Title Video Camera Based Monitoring System

Abstract



OBJECT: To automatically recognise an object present in a monitoring region and to inform a monitoring person by voice, and to automatically track a noteworthy object.

CONSTITUTION: A video camera based monitoring system comprising: a means for obtaining a wide-angle image of an entire monitoring region; a means for processing this wide-angle image, determining a noteworthy object and obtaining its position and size; a means for moving the line-of-sight of a camera to this noteworthy object and obtaining a close-up image of the object; a means for processing this close-up image and recognising the object; and a means for providing a voice-based notification of the recognition result; whereby an object present in the monitoring region is automatically recognised, notification is given by voice, and the noteworthy object is automatically tracked.

EFFECT: The invention eliminates the need for uninterrupted monitoring of a monitor screen and therefore saves labour. Because a close-up image is constantly obtained, no time has to be spent in searching for an abnormal body in an emergency.

Claim

1. A video camera based monitoring system characterised in that it comprises:

a means for obtaining a wide-angle image of an entire monitoring region;

a means for processing this wide-angle image, determining a noteworthy object and obtaining its position and size;

a means for moving the line-of-sight of a camera to this noteworthy object and obtaining a close-up image of the object;

a means for processing this close-up image and recognising the object; and

a means for providing a voice-based notification of the recognition result.

Detailed Description of the Invention

Industrial field of utilisation

(1) The present invention relates to monitoring systems, and in particular to technology which is effective in application to video camera based monitoring systems.*

Prior art

(2) Hitherto, video camera based monitoring has involved a person looking at and monitoring an image obtained from a video camera, and on this basis making a decision regarding the state of, or situation in, a particular region. The following three methods have been employed in this video camera based monitoring system:

(A) surveying an entire monitoring region at one time by means of a wide-angle fixed camera;

(B) surveying an entire monitoring region in a fixed period of time by periodically shifting the line-of-sight of a camera with a standard angle of field;

(C) monitoring a target region by means of a monitoring person remotely controlling a camera with a variable angle of field and a movable line-of-sight.

Problems that the invention will solve

(3) However, in each of these methods, the basis of the monitoring is a person looking at an image from a video camera. Consequently, unless used in conjunction with another sensor system which detects a change or abnormality, most of the time it is necessary to keep on looking at an unchanging image, which is wasted effort.

(4) Moreover, even if, by concurrent use of another sensor system, the image is viewed only when a change has occurred, the following problems have been encountered with the conventional systems noted above:

(i) With system (A), because a wide-angle camera is used, the image of an individual object is small and it may be impossible for a viewing person to determine the state of the object.

(ii) With system (B), it usually takes a certain length of time before the line-of-sight of the camera is directed towards an object that has undergone a change, and then the line-of-sight of the camera immediately diverges away from the object.

(iii) With system (C), as in system (B), a certain time is required for the line-of-sight of the camera to come into alignment with an object that has undergone a change. In addition, it takes time to adjust the camera to an angle of field suited to the size of the object.

(5) It is an object of the present invention to provide a technique whereby the waste of effort incurred in uninterruptedly monitoring a monitor screen can be eliminated, a noteworthy object located within a monitoring region can be automatically tracked, and time need not be spent searching for an abnormal body in the event of an emergency.

(6) These objects of the present invention, together with further objects and novel features, will be made clear by this detailed description of the invention and by the accompanying drawings.

---

* Numbers in round brackets at the beginning of paragraphs correspond to the paragraph numbering in the Japanese patent document.

Means for solving problems

(7) In order to attain the above-described objects, the present invention is a video camera based monitoring system whereof the principal distinguishing feature is that it comprises: a means for obtaining a wide-angle image of an entire monitoring region; a means for processing this wide-angle image, determining a noteworthy object and obtaining its position and size; a means for moving the line-of-sight of a camera to this noteworthy object and obtaining a close-up image of the object; a means for processing this close-up image and recognising the object; and a means for providing a voice-based notification of the recognition result.

Working of the invention

(8) Given the provision of the various means described above, the system of the invention operates as follows. (i) The wide field angle fixed camera is used to constantly acquire a wide-angle image of an entire monitoring region. (ii) This wide-angle image is processed, a noteworthy object is extracted, and its position and size are obtained. (iii) The variable field angle, movable line-of-sight camera is used, its line-of-sight aligned with the aforementioned object, and its angle of field adjusted, thereby obtaining a close-up image of the object. (iv) This close-up image is processed and the identity of the object is recognised. (v) This recognition result is conveyed to a monitoring person by a voice message.

(9) The above working makes it possible to automatically recognise an object present within a monitoring region, to notify a monitoring person by means of a voice message, and to automatically track a noteworthy object.

(10) The constitution of the present invention will now be described with reference to an embodiment.

(11) Note that in all the drawings serving to clarify the embodiment, elements having an identical function are given identical referencing numerals, and such elements are not repeatedly described.

Embodiment

(12) FIG. 1 is a block diagram of an embodiment of the present invention. FIG. 2 shows an example of an output image of the video combiner of FIG. 1. FIG. 3 shows the processing flow in the first image processor of FIG. 1. FIGS. 4A–4F show, in each of six temporally successive frames, the result of outline extraction of an input image. FIGS. 5A–5F show, in six corresponding frames, the result of log-polar coordinate transformation of each frame of FIGS. 4A-4F. FIGS. 6A–6F show, in corresponding frames, the inspection images [1]* derived for each frame of FIGS. 4A-4F. FIG. 7 shows the vertical and horizontal pixel frequencies in the inspection image of the first frame, shown in FIG. 6A. FIG. 8 shows the processing flow in the second image processor of FIG. 1.

(13) As indicated in FIG. 1, the video camera based monitoring system of the present embodiment firstly obtains an image of an entire monitoring region with wide field angle fixed camera 11. This wide field angle fixed camera 11 can be an ordinary video camera fitted with a wide-angle lens. The image obtained is input to first image processor 12. First image processor 12 processes the image and outputs the position

* Numbers in square brackets refer to Translator's Notes appended to the translation.

and size of a noteworthy object, this being a body present within the monitoring region. The processing performed by first image processor 12 will be described hereinafter. It may be noted that "noteworthy object" indicates something whose identity is unclear and something which is moving, as in the case of something which has appeared for the first time in the monitoring region.

(14) The output of first image processor 12 is sent to variable field angle movable camera 13. This is a video camera fitted with a motorised zoom lens and attached to a tripod head providing motor-controllable vertical and horizontal rotation. Variable field angle movable camera 13 obtains a close-up image of a target object by controlling the tripod head and the zoom lens on the basis of position and size information in the output of first image processor 12. The close-up image is sent to second image processor 14 which performs recognition of the target object. The processing performed by second image processor 14 will be described hereinafter.

(15) The result of the processing performed by second image processor 14 is conveyed to a monitoring person by a voice message produced by speech synthesizer 15 and speaker 16. The image from wide field angle fixed camera 11, the image from variable field angle movable camera 13, and the target object recognition result output from second image processor 14, are input to video combiner 17 which combines the images for purposes of monitoring. This combined image is displayed on video monitor 18 where it can be watched by a monitoring person. Note that video combiner 17 and video monitor 18 are provided if the need arises, and are not elements that are essential to the present invention.

(16) An example of a combined image for monitoring purposes is shown in FIG. 2, where it will be seen that screen 21 of video monitor 18 displays images 22 to 26 of objects within the monitoring region, and also displays the line-of-sight position 27 of the close-up image. Line-of-sight position 27 is superimposed on image 22 of a noteworthy object, and close-up image 28 of image 22 of this object is displayed in a portion 29 of the screen. Note that this close-up image 28 is not displayed when there is no noteworthy object or body. In such a case, line-of-sight position 27 is located in the centre of the screen.

(17) Next, the processing flow in first image processor 12 will be described with reference to FIGS. 3 to 7.

(18) The processing flow in first image processor 12 will first of all be described in accordance with the flowchart of FIG. 3.

(19) The following processing is something which is repeated for each frame of an input image.

(20) At step 301, the wide-angle image from wide field angle fixed camera 11 is input.

(21) At step 302, outlines of bodies are extracted from the input wide-angle image. Many techniques have been proposed for this, and for example the one disclosed in Y. Hongo, M. Kawahito, T. Inui and M. Miyake, "Outline Extraction by a Locally Parallel Stochastic Algorithm with Energy Learning Function", *Trans. IEICE of Japan*, D-II, Vol. J74-D-II, No.3, pp.348–356 (1991) can be utilised. The outline extraction provides the outline images 4a–4f shown in FIGS. 4A–4F. The following line figures are

displayed in these outline images: a circle 41 (hereinafter referred to as ○ 41), a triangle 42 (hereinafter referred to as ρ 42) and a square 43 (hereinafter referred to as □ 43). Line-of-sight position 27 is also displayed in the outline images.

(22) At step 303, a labelling process is performed, this labelling applying numbers to each line figure representing the outline of a body. The technique given in J. Hasegawa, Y. Koshimizu, A. Nakayama and S. Yokoi, "Fundamental Techniques of Image Processing: Introduction to Techniques", pub. Gijutsu Hyoronsha, pp.45–49 (1986) can be utilised for this labelling. As a result of this processing, the number 1 is assigned to ○ 41, the number 2 to ρ 42, and the number 3 to □ 43 of FIG. 4A.

(23) At step 304, log-polar coordinate transformation is performed. Log-polar coordinate transformation transforms coordinates in a space defined by plotting $\log(\boldsymbol{r})$ along the horizontal axis and $\boldsymbol{\theta}$ along the vertical axis, where $\boldsymbol{r}$ is the distance from the centre of an image and $\boldsymbol{\theta}$ is the angle of rotation from the horizontal axis. Note that this transformation is not essential to the present invention. However, as shown in I. Yoroisawa, "Image Sampling Method", Jap. Pat. Appl. No. 03-097032 (1991), performing this transformation enables the amount of computation involved in subsequent image processing to be greatly decreased. The ability to monitor in real time is an important property of a monitoring system and hence there is a need to reduce the amount of computation as much as possible. This transformation is therefore adopted in the present embodiment. Transformed images 5a–5f obtained as a result of this transformation are shown in FIGS. 5A–5F. Transformed line figures 51, 52 and 53 are displayed in these transformed images.

(24) At step 305, a difference image is obtained by computing (transformed image)–(pre-transformation image) and changing negative pixel values to 0. Here, "pre-transformation image" means an immediately (i.e., by 1 frame) prior transformed image. The pre-transformation image in the case of the first frame is taken as 0 (meaning that all pixel values are 0).

(25) At step 306, it is decided whether or not this difference equals 0. Provided that the body has not disappeared from the screen, a difference equal to 0 means that there has been no change in the image. If the difference is not 0, processing advances to step 307, while if the difference is 0 it advances to step 308.

(26) At step 307, the difference image is set as the inspection image. Inspection images 6a–6f obtained in this embodiment are shown in FIGS. 6A–6F. Line figures 61, 62 and 63 are displayed in these inspection images.

(27) At step 308, the stored image is set as the inspection image. Here, the "stored image" is the image obtained by removing, at each successive frame, from the initial transformed image, a body that has become a noteworthy object. The stored image in respect of the first frame is taken as transformed image 5a.

(28) At step 309, the vertical and horizontal pixel frequencies of the inspection image obtained in step 307 or step 308 are counted. The results of such counting in the case of the first frame are shown in FIG. 7.

(29) At step 310, it is decided whether or not both the vertical and horizontal pixel frequencies are 0. If these pixel frequencies are not 0, processing advances to step 311, while if they are 0 processing advances to step 314.

(30) At step 311, the positions of maximum vertical and horizontal frequency are extracted.

(31) At step 312, the centre of the input image body outline [2] having the label of the pixel which is present at the position of maximum vertical and horizontal frequency or at the position nearest to this position, is extracted as the line-of-sight position.

(32) At step 313, the size of the target body is measured.

(33) At step 314, line-of-sight position 27 is set as the centre of the screen and the size of the object is set to 0.

(34) At step 315, the line-of-sight position and the size of the target body are output. Variable field angle movable camera 13 obtains a close-up image of the target by controlling the tripod head and the zoom lens on the basis of this output.

(35) At step 316, (stored image)−(target body outline) is set as the new stored image and stored.

(36) The processing described above is repeated for each frame of the input image.

(37) Next, the processing set out in FIG. 3 will be described in detail for each frame of an input image.

(38) FIGS. 4A–4F show, in each of six temporally successive frames, the results of the outline extraction of step 302 being applied to an input image. FIGS. 5A–5F show, in each of the frames, the results of the log-polar coordinate transformation of step 304. FIGS. 6A–6F show, in each of the frames, the inspection images obtained by step 307 or step 308. Note that the size of the wide-angle image is 64 ● 64 pixels.

(39) The nature of the frame-by-frame changes in the results of the outline extraction shown in FIGS. 4A–4F will now be described.

(40) The first frame, shown in FIG. 4A, is the initial frame. Line figures ○ 41, ρ 42 and □ 43 are displayed in the image 4a of this frame, these line figures being body outlines. The image 4b of the second frame is exactly the same as image 4a of the first frame. In the image 4c of the third frame, the position of □ 43 has changed. The image 4d of the fourth frame is exactly the same as image 4c of the third frame. The image 4e of the fifth frame is exactly the same as image 4d of the fourth frame. In the image 4f of the sixth frame, the position of □ 43 has changed.

(41) Firstly, the processing in the case of the first frame will be described. Because the first frame is the initially appearing frame, transformed image 5a is used as the stored image and the pre-transformation image is taken as 0. Accordingly, the difference image in step 305 is simply the transformed image 5a itself. As a result, because the difference is not 0, processing advances from step 306 to step 307 and the difference image is taken as inspection image 6a. In inspection image 6a, line figures 61, 62 and 63 are displayed as noteworthy objects. As shown in FIG. 7, the vertical and horizontal pixel frequencies of this inspection image 6a are maximum at vertical = 2, horizontal = 44. The label of the pixel at this position is 3 and in input image 4a the centre of body outline □ 43, which has this label, is the position given by vertical = 27, horizontal = 52. The size of this body outline □ 43 (the length of one side of a circumscribed square) is 12. This position and size are output, whereby □ 43

is taken as the noteworthy object and becomes a close-up image. [3] A new stored image is then obtained by removing these label 3 pixels from the stored image.

(42) Next, in the second frame, because transformed image 5b is exactly the same as pre-transformation image 5a, the difference image at step 306 becomes 0. As a result, the stored image is used as inspection image 6b. In inspection image 6b, line figures 61 and 62 corresponding to O 41 and ρ 42 are displayed. Although these line figures correspond to noteworthy objects, they have not yet become targets for close-ups. The vertical and horizontal pixel frequencies of this inspection image 6b are maximum at vertical = 38, horizontal = 54. The label of the pixel at this position is 1 and in the input image the centre of body outline O 41, which has this label, is the position given by vertical = 14, horizontal = 17. The size of this body outline is 18. This position and size are output, whereby O 41 is taken as the noteworthy object and becomes a close-up image. A new stored image is then obtained by removing these label 1 pixels from the stored image.

(43) Next, in the third frame, because the position of □ 43 in image 4c has changed, transformed image 5c is different from pre-transformation image 5b and hence the difference image is not 0. As a result, the difference image is used as inspection image 6c. In inspection image 6c, line figure 63 corresponding to □ 43 with altered position is displayed. The vertical and horizontal pixel frequencies of this inspection image 6c are maximum at vertical = 64, horizontal = 54. The label of the pixel at this position is 3 and in the input image the centre of body outline □ 43, which has this label, is the position given by vertical = 25, horizontal = 50. The size of this body outline is 12. This position and size are output, whereby □ 43 is taken as the noteworthy object and becomes a close-up image. A new stored image is then obtained by removing these label 3 pixels from the stored image. Note that because label 3 pixels have already been removed in the processing of the first frame, the stored image is effectively not updated.

(44) Next, in the fourth frame, because transformed image 4d is exactly the same as pre-transformation image 4c [4], the difference image becomes 0. As a result, the stored image is used as inspection image 6d. In inspection image 6d, line figure 62 [5] corresponding to ρ 42 is displayed. Although this line figure corresponds to an unknown object, it has not yet become a target for a close-up. The vertical and horizontal pixel frequencies of this inspection image 6d are maximum at vertical = 26, horizontal = 52. The label of the pixel at this position is 2 and in the input image the centre of body outline ρ 42, which has this label, is the position given by vertical = 46, horizontal = 15. The size of this body outline is 13. This position and size are output, whereby ρ 42 is taken as the noteworthy object and becomes a close-up image. A new stored image is then obtained by removing label 2 pixels from the stored image. At this point in the processing, the stored image becomes 0.

(45) Next, in the fifth frame, transformed image 5e is exactly the same as pre-transformation image 5d and hence the difference image is 0. In this case, the stored image is used as inspection image 6e. The vertical and horizontal pixel frequencies of this inspection image 6e are both 0. Consequently, the line-of-sight position is the centre of the screen (vertical = 32, horizontal = 32) and the size of the body is 0. These values are output. This output signifies that there is no target body for a close-up image. Hence the stored image is effectively not updated.

(46) Next, in the sixth frame, because the position of □ 43 in image 4f has changed, transformed image 5f is different from pre-transformation image 5e and hence the difference image is not 0. In this case, the difference image is used as inspection image 6f. In inspection image 6f, line figure 63 corresponding to □ 43 with altered position is displayed. The vertical and horizontal pixel frequencies of this inspection image are maximum at vertical = 2, horizontal = 44. The label of the pixel at this position is 3 and in the input image the centre of body outline □ 43, which has this label, is the position given by vertical = 27, horizontal = 52. The size of this body outline is 12. This position and size are output, whereby □ 43 is taken as the noteworthy object and becomes a close-up image. The stored image is effectively not updated.

(47) Similar processing is performed for each subsequent frame, and when an unknown object appears for the first time within the monitoring region, or an object has moved within the monitoring region, the relevant information is output in order to obtain a close-up image.

(48) Next, the processing flow in second image processor 14 of FIG. 1 will be described with reference to FIG. 8. This second image processor 14 operates on each output frame of variable field angle movable camera 13 when the size of the object is not 0. Most of this processing is the same as that used in I. Yoroisawa, "Drawing Interpretation Processing System", Jap. Pat. Appl. No. 03-021623 (1991).

(49) At step 801, close-up image 28 from variable field angle movable camera 13 is input, the line-of-sight direction and field angle of this camera 13 having been controlled on the basis of the output of first image processor 12.

(50) At step 802, the outline of the body is extracted from the input close-up image. For this, it is convenient to use a similar technique to that employed in step 302 of FIG. 3.

(51) At step 803, log-polar coordinate transformation is performed, and rotations and changes in the apparent size of the target body are transformed into translations. This technique is the same as that used in step 304 of FIG. 3.

(52) At step 804, the R-transformation is performed, yielding an image which is invariant under translation. For this R-transformation, the principles disclosed in H. Reitboeck and T. P. Brody, "A transformation with invariance under cyclic permutation for applications in pattern recognition", *Information and Control*, 15(2), 130–154 (1969), can be utilised.

(53) At step 805, the result of the R-transformation is compared with dictionary images.

(54) At step 806, the name of the dictionary image for which the difference (the vector distance) is smallest is identified as the name of the target body. Recognition is achieved by means of this processing.

(55) Finally, at step 807, the recognition result is output. This target body recognition result is then conveyed to a monitoring person by means of a voice message, as described previously.

(56) The present invention has now been described in specific detail on the basis of the foregoing embodiment. However, the invention is not limited to this embodiment

and can of course be modified in various ways within a scope that does not depart from the spirit of the invention.

Effect of the invention

(57) As has been described above, the present invention is capable of providing automatic recognition of a body located in a monitoring region, and communication of this recognition result by means of a voice message. This eliminates the need for uninterrupted monitoring of a monitor screen and therefore saves labour.

(58) The invention also makes it possible to automatically track a noteworthy object — i.e., something located in the monitoring region whose identity is unclear and something which is moving. Because a close-up image of this is constantly obtained, no time has to be spent in searching for an abnormal body in an emergency.

(59) The present invention will be effective in saving labour in industries where monitoring is required, such monitoring including for example security of buildings and parks, and surveys of deep-sea regions and regions of radioactive contamination.

Brief Description of the Drawings

FIG. 1 is a block diagram of an embodiment of the present invention.

FIG. 2 shows an output image from the video combiner of FIG. 1.

FIG. 3 is a flowchart showing the processing flow in the first image processor of FIG. 1.

FIGS. 4A–4F respectively show the first to the sixth frames of the result of the input image outline extraction performed at step 302 of FIG. 3.

FIGS. 5A–5F respectively show the first to the sixth frames of the result of the log-polar coordinate transformation performed at step 304 of FIG. 3.

FIGS. 6A–6F respectively show the first to the sixth frames of the inspection image obtained at step 307 or 308 of FIG. 3.

FIG. 7 shows the vertical and horizontal pixel frequencies of FIG. 6A.

FIG. 8 shows the processing flow in the second image processor of FIG. 1.

Explanation of referencing numerals

11 ....wide field angle fixed camera
12 ....first image processor
13 ....variable field angle movable camera
14 ....second image processor
15 ....voice synthesizer
16 ....speaker
17 ....video combiner
18 ....video monitor
21 ....screen of video monitor
27 ....line-of-sight position of close-up image
28 ....close-up image

FIG. 1

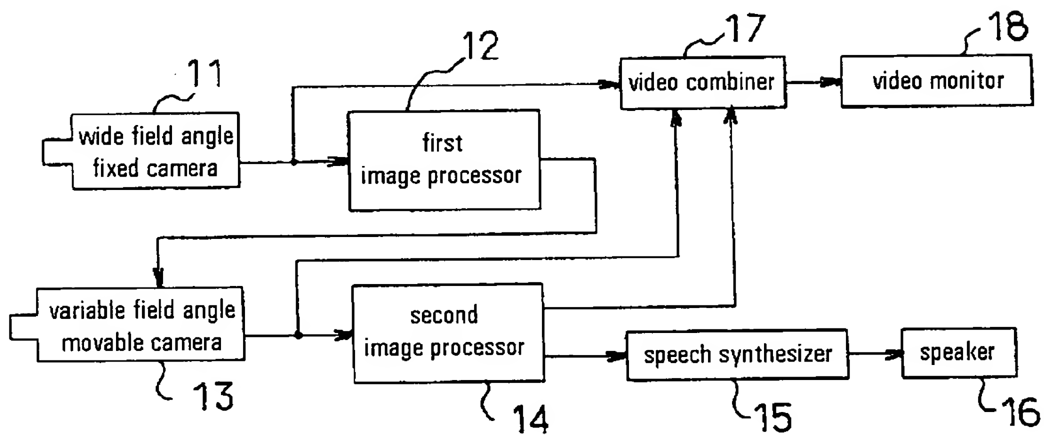


FIG. 2

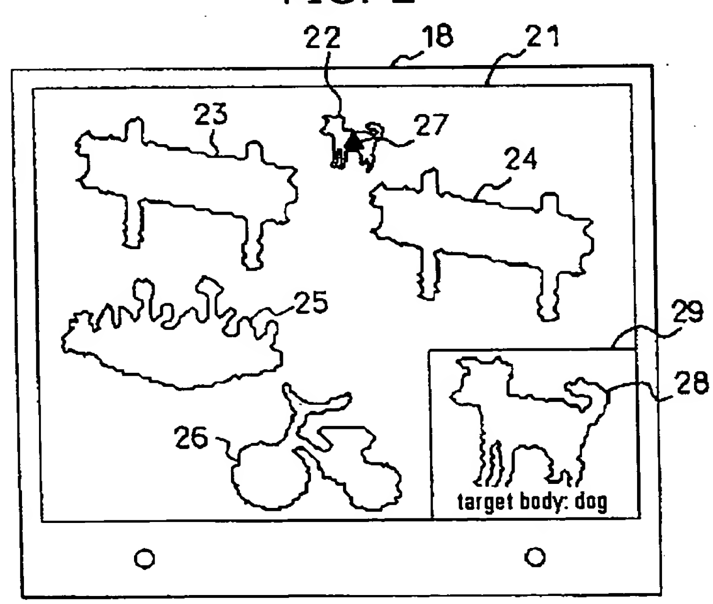


FIG. 3

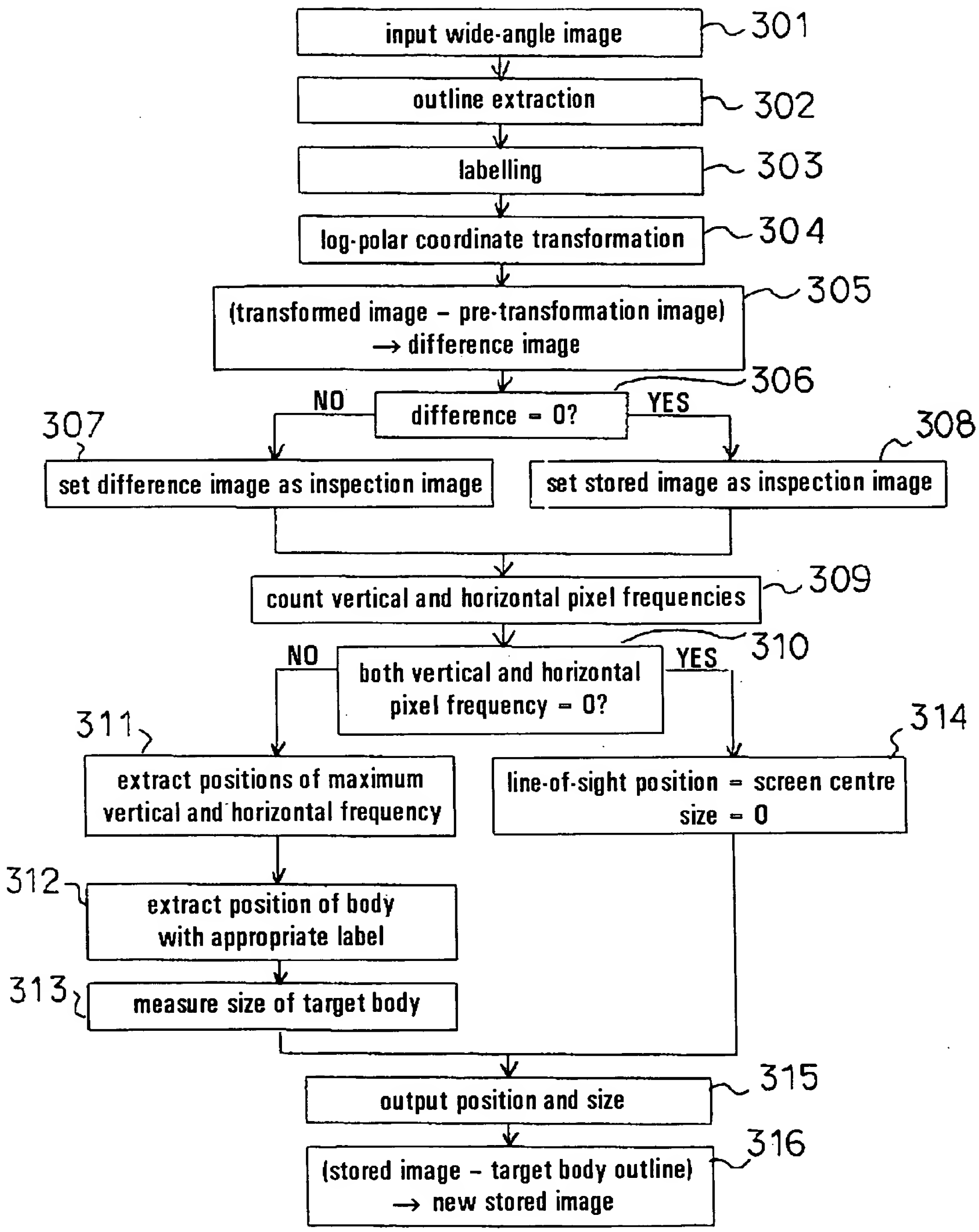

input wide-angle image
301
outline extraction
302
labelling
303
log-polar coordinate transformation
304
(transformed image – pre-transformation image) → difference image
305
difference = 0?
306
NO
YES
set difference image as inspection image
307
set stored image as inspection image
308
count vertical and horizontal pixel frequencies
309
both vertical and horizontal pixel frequency = 0?
310
NO
YES
extract positions of maximum vertical and horizontal frequency
311
line-of-sight position = screen centre
size = 0
314
extract position of body with appropriate label
312
measure size of target body
313
output position and size
315
(stored image – target body outline) → new stored image
316

FIG. 4

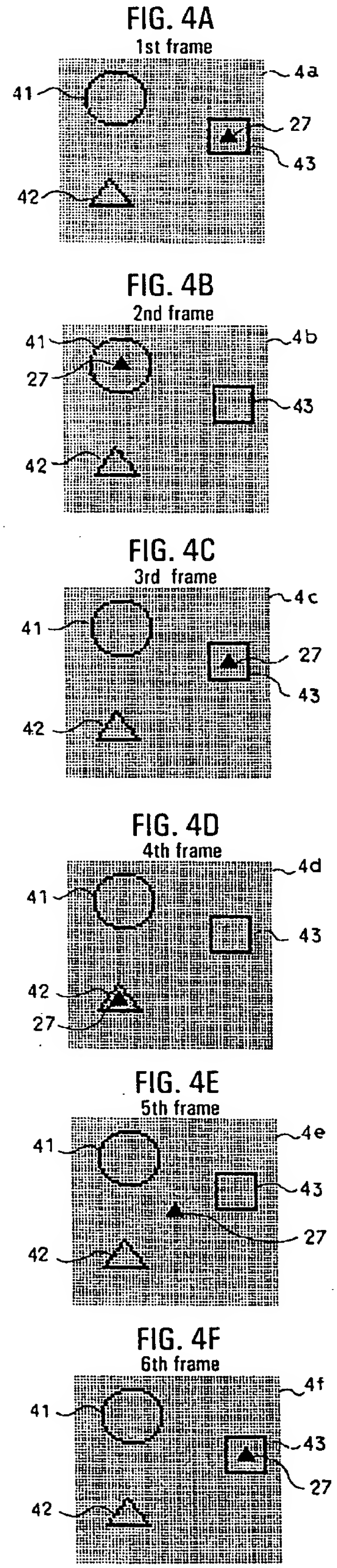


FIG. 5

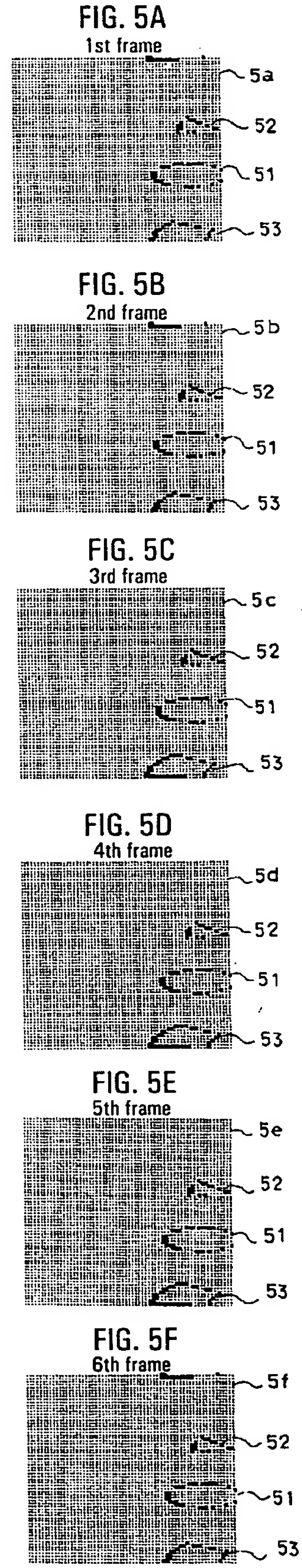


FIG. 6

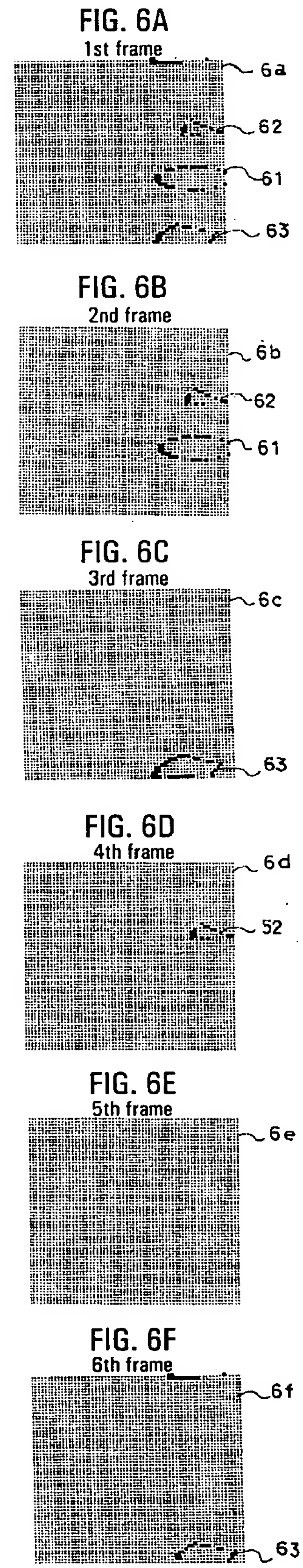


FIG. 7

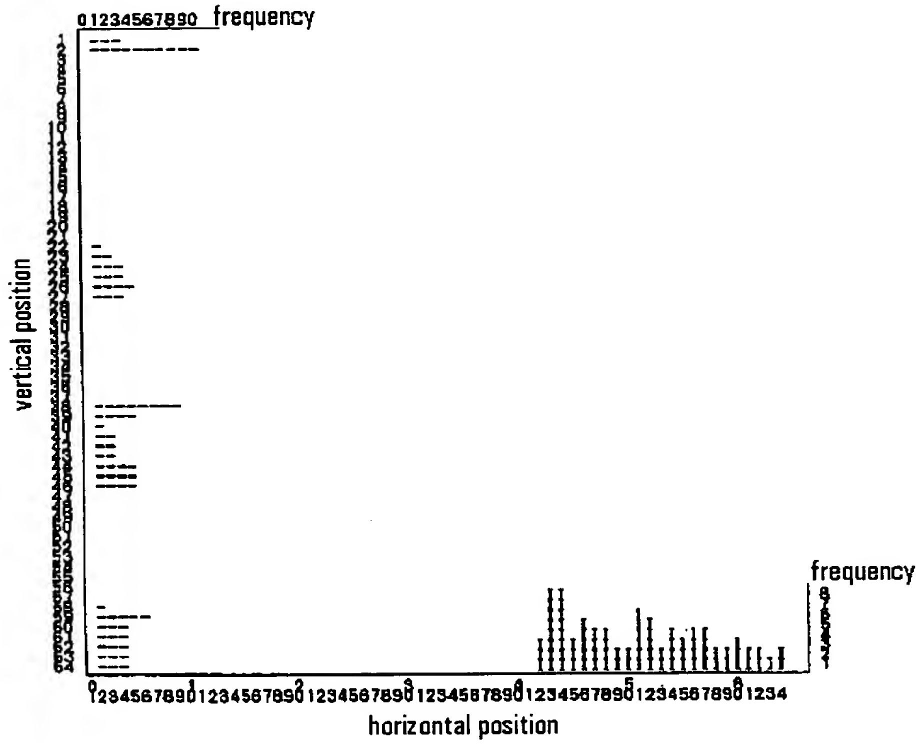
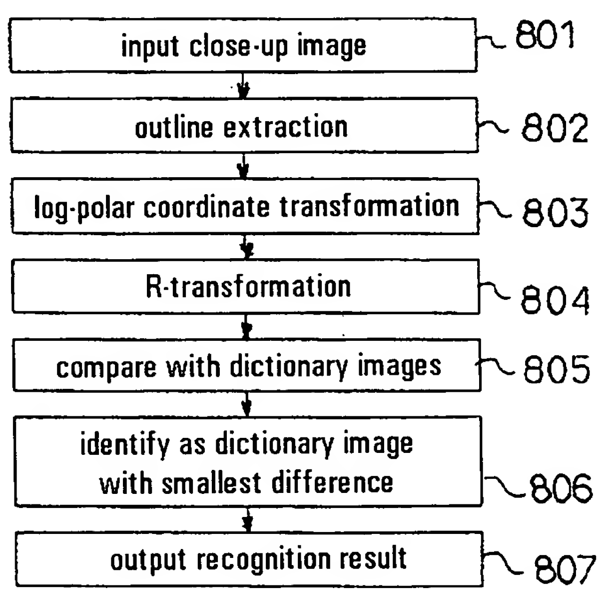

frequency
vertical position
frequency
horizontal position

FIG. 8


input close-up image
801
outline extraction
802
log-polar coordinate transformation
803
R-transformation
804
compare with dictionary images
805
identify as dictionary image
with smallest difference
806
output recognition result
807

TRANSLATOR'S NOTES

1. Although the literal meaning of the Japanese term that I have translated as "inspection images" is quite clear, the actual meaning of this term in the present context is not clear to me.
2. By "the centre of the input image body outline", the writer presumably means the centre of an outline (of a body) derived, by outline extraction, from the input image of a monitoring region.
3. Sic. The writer presumably means that □ 43 is taken as the noteworthy object and a close-up image is obtained of the body of which □ 43 is the extracted outline.
4. Sic. However, images 4c and 4d are properly speaking outline images, not images after the log-polar coordinate transformation. As far as I can see, the correct terminology here would be "because transformed image ***5d*** is exactly the same as pre-transformation image ***5c***".
5. Note that FIG. 6D in the Japanese document erroneously labels this line figure as "52".